DOCUMENT RESUME

ED 482 122 IR 058 806

AUTHOR	Milekic, Slavko
TITLE	The More You Look the More You Get: Intention-Based Interface Using Gaze-Tracking.
PUB DATE	2003-00-00
NOTE	20p.; In: Museums and the Web 2003: Selected Papers from an International Conference (7th, Charlotte, NC, March 19-22, 2003); see IR 058 801.
AVAILABLE FROM	For full text: http://www.archimuse.com/mw2003/papers/milekic/milekic.html/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Interfaces; Eye Contact; *Eye Movements; Eyes; *Intention; Visual Measures; *Visual Perception
IDENTIFIERS	Digital Technology; *Gaze Patterns; *Visual Tracking

ABSTRACT

Only a decade ago eye- and gaze-tracking technologies using cumbersome and expensive equipment were confined to university research labs. However, rapid technological advancements (increased processor speed, advanced digital video processing) and mass production have both lowered the cost and dramatically increased the efficacy of eye- and gaze-tracking equipment. This opens up a whole new area of interaction mechanisms with museum content. This paper describes a conceptual framework for an interface, designed for use in museums and galleries, which is based on non-invasive tracking of a viewer's gaze direction. Following the simple premise that prolonged visual fixation is an indication of a viewer's interest, the author dubbed this approach intention-based interface. Includes 16 figures. (Contains 27 references.) (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 482 122

PAPERS
Museums and the Web 2003

The More You Look The More You Get: Intention-Based Interface Using Gaze-Tracking

Slavko Milekic, The University of the Arts, USA

http://www.uarts.edu

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Bearman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
□ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Register
Workshops
Sessions
Speakers
Interactions
Demonstrations
Exhibits
Events
Best of the Web
Key Dates
Charlotte

A&MI

Archives & Museum Informatics
158 Lee Avenue
Toronto Ontario
M4E 2P3 Canada

ph: +1 416-691-2516
fx: +1 416-352-6025

info@archimuse.com
www.archimuse.com

Search A&MI

Join our Mailing List.
Privacy.

Updated: March 13, 2003

Abstract

Only a decade ago eye- and gaze-tracking technologies using cumbersome and expensive equipment were confined to university research labs. However, rapid technological advancements (increased processor speed, advanced digital video processing) and mass production have both lowered the cost and dramatically increased the efficacy of eye- and gaze-tracking equipment. This opens up a whole new area of interaction mechanisms with museum content. In this paper I will describe a conceptual framework for an interface, designed for use in museums and galleries, which is based on non-invasive tracking of a viewer's gaze direction. Following the simple premise that prolonged visual fixation is an indication of a viewer's interest, I dubbed this approach intention-based interface.

Keywords: eye tracking, gaze tracking, intention-based interface

Introduction

In humans, gaze direction is probably the oldest and earliest *means of communication at a distance*. Parents of young infants are often trying to 'decode' from an infant's gaze direction the needs and interest of their child. Thus, gaze direction can be viewed as a first instance of *pointing*. A number of developmental studies (Scaife and Bruner 1975; Corkum and Moore, 1988; Moore 1999) show that even very young infants actively follow and respond to the gaze direction of their caregivers. The biological significance of eye movements and gaze direction in humans is illustrated by the fact that humans, unlike other primates, have visible white area (sclera) around the pigmented part of the eye (iris, covered by transparent cornea, see Figure 1). This makes even discrete shifts of gaze direction very noticeable (as is painfully obvious in cases of 'lazy eye').

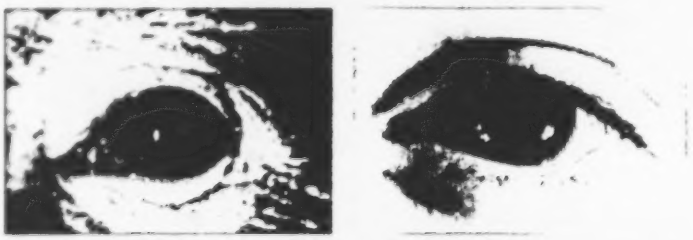

Figure 1. Comparison of human and non-human eye (chimpanzee). Although many animals have pigmentation that accentuates the eyes, the visible white area of human eye makes it easier to interpret the gaze direction

Eye contact is one of the first behaviors to develop in young infants. Within the first few days of life, infants are capable of focusing on their caregiver's eyes (Infants are

IR058806

2 BEST COPY AVAILABLE

physiologically shortsighted with the ideal focusing distance of 25-40 cm. This distance corresponds to the distance between the mother's and infant's eyes when the baby is held at the breast level. Everything else is conveniently a blur. Within the first few weeks, establishing eye contact with the caregiver produces a smiling reaction (Stewart & Logan, 1998). Eye contact and gaze direction continue to play a significant role in social communication throughout life. Examples include:

- regulating conversation flow;
- regulating intimacy levels;
- indicating interest or disinterest;
- seeking feedback;
- expressing emotions;
- influencing;
- signaling and regulating social hierarchy;
- indicating submissiveness or dominance;

Thus, it is safe to assume that humans have a large number of behaviors associated with eye movements and gaze direction. Some of these are innate (orientation reflex, social regulation), and some are learned (extracting information from printed text, interpreting traffic signs).

Our relationship with works of art is essentially a social and intimate one. In the context of designing a gaze tracking-based interface with cultural heritage information, innate visual behaviors may play a significant role precisely because they are social and emotional in nature and have the potential to *elicit a reaction* external to the viewer. In this paper I will provide a conceptual framework for the design of gaze-based interactions with cultural heritage information using the digital medium. Before we proceed, it is necessary to clarify some of the basic physiological and technological terms related to eye- and gaze-tracking.

Eye Movements and Visual Perception

While we are observing the world, our subjective experience is that of a smooth, uninterrupted flow of information and a sense of the wholeness of the visual field. This, however, contrasts sharply with what actually happens during visual perception. Our eyes are stable only for brief periods of time (200-300 milliseconds) called *fixations*. Fixations are interspersed by rapid, jerky movements called *saccades*. During these movements no new visual information is acquired. Furthermore, the information gained during the periods of fixations is clear and detailed only in a small area of the visual field - about 2° of visual angle. Practically, this corresponds to the area covered by one's thumb at arm's length. The rest of the visual field is fuzzy but provides enough information for the brain to plan the location of the next fixation point. The problems that arise because of the discrepancy between our subjective experience and the data gained by using eye-tracking techniques can be illustrated by the following example:

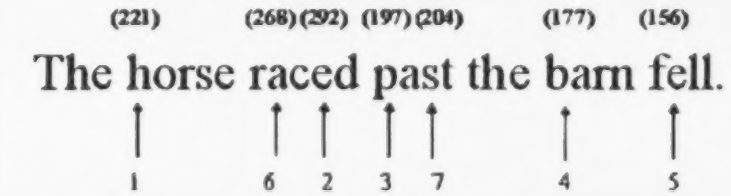

Figure 1

The sentence above is a classical example of a "garden path" sentence that (as you probably have experienced) initially leads the reader to a wrong interpretation (Bever, 1970). The eye-tracking data provide information about the sequence of fixations (numbered 1 to 7) and their duration in milliseconds. The data above provide some clues about the relationship between visual analysis during reading and eye movements. For example, notice the presence of two *retrograde saccades*

(numbered 6 and 7) that happened after initial reading of the sentence. They more than double the *total fixation time* of the part of the sentence necessary for disambiguation of its meaning. Nowadays there is a general consensus in the eye-tracking community that the number and the duration of fixations are related to the cognitive load imposed during visual analysis.

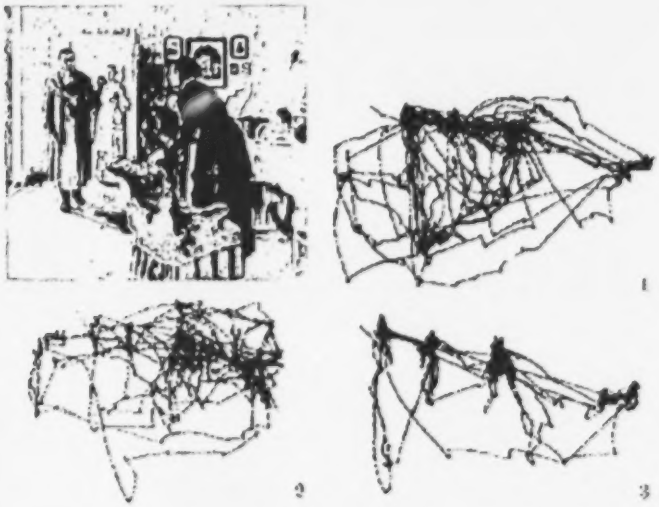

Figure 2. Illustration of differences in gaze paths while interpreting I. Repin's painting "They did not expect him."

Path (1) corresponds to free exploration. Path (2) was obtained when subjects were asked to judge the material status of the family, and path (3) when they were asked to guess the age of different individuals. Partially reproduced from Yarbus, A. L. (1967)

Eye-tracking studies of reading are very complex but have the advantage of allowing fine control of different aspects of the visual stimuli (complexity, length, exposure time, etc.). Interpretation of eye movement data during scene analysis is more complicated because visual exploration strategy is heavily dependent on the *context of exploration*. Data (Figure 2) from an often-cited study by Yarbus (1967) illustrate differences in visual exploration paths during interpretation of Ilya Repin's painting "They did not expect him, or "the unexpected guest".

Brief History of Eye- and Gaze-Tracking

The history of documented eye- and gaze-tracking studies is over a hundred years old (Javal, 1878). It is a history of technological and theoretical advances where progress in either area would influence the other, often producing a burst of research activity that would subsequently subside due to the uncovering of a host of new problems associated with the practical uses of eye-tracking.

Not surprisingly, the first eye-tracking studies used other humans as tracking instruments by utilizing strategically positioned mirrors to infer gaze direction. Experienced psychotherapists (and socially adept individuals) still use this technique, which, however imperfect it may seem, may yield a surprising amount of useful information. Advancements in photography led to the development of a technique based on capturing the light reflected from the cornea on photographic plate (Dodge & Cline, 1901). Some of these techniques were fairly invasive, requiring placement of a reflective white dot directly onto the eye of the viewer (Jud, McAllister & Steel, 1905) or a tiny mirror, attached to the eye with a small suction cup (Yarbus, 1967). In the field of medicine a technique was developed (electro-oculography, still in use for certain diagnostic procedures) that allowed registering of eyeball movements using a number of electrodes positioned around the eye. Most of the described techniques

required the viewer's head to be motionless during eye tracking and used a variety of devices like chin rests, head straps and bite-bars to constrain the head movements. The major innovation in eye tracking was the invention of a head-mounted eye tracker (Hartridge & Thompson, 1948). With technological advances that reduced the weight and size of an eye tracker to that of a laptop computer, this technique is still widely used.

Most eye tracking techniques developed before the 1970s were further constrained by the fact that data analysis was possible only *after* the act of viewing. It was the advent of mini- and microcomputers that made possible real-time eye tracking. Although widely used in studies of perceptual and cognitive processes, it was only with the proliferation of personal computers in the 1980s that eye tracking was applied as an instrument for the evaluation of human-computer interaction (Card, 1984). Around the same time, the first proposals for the use of eye tracking as a means for user-computer communication appeared, focusing mostly on users with special needs (Hutchinson, 1989; Levine, 1981). Promoted by rapid technological advancements, this trend continued, and in the past decade a substantial amount of effort and money was devoted to the development of eye- and gaze-tracking mechanisms for human-computer interaction (Vertegaal, 1999;Jacob, 1991; Zhai, Morimoto & Ihde, 1999). Detailed analysis of these studies is beyond the scope of this paper, and I will refer to them only insofar as they provide reference points to my proposed design. Interested readers are encouraged to consult several excellent publications that deal with the topic in much greater detail (Duchowsky, 2002; Jacob, Karn, 2003 /in press/).

Eye and Gaze Tracking in a Museum Context

The use of eye and gaze tracking in a museum context extends beyond interactions with the digital medium. Eye tracking data can prove to be extremely useful in revealing how humans observe real artifacts in a museum setting. The sample data and the methodology from a recent experiment conducted in the National Gallery in London (in conjunction with the Institute for Behavioural Studies) can be seen on the Web. Although some of my proposed gaze-based interaction solutions can be applied to the viewing of real artifacts (for example, to get more information about particular detail that a viewer is interested in), the main focus of my discussion will be on the development of affordable and intuitive gaze-based interaction mechanisms with(in) the digital medium. The main reason for this decision is the issue of accessibility to cultural heritage information. Although an impressive 4000 people participated in the National Gallery experiment, they *all had to be there* at certain time. I am not disputing the value of experiencing the real artifact, but the introduction of the digital medium has dramatically shifted the role of museums from *collection & preservation* to *dissemination & exploration*. Recent advancements in Web-based technologies make it possible for museums to develop tools (and social contexts) that allow them to serve as centers of knowledge transfer for both local and virtual communities. My proposal will focus on three issues:

1. problems associated with use of gaze tracking data as interaction mechanism;
2. conceptual framework for the development of gaze-based interface;
3. currently existing (and affordable) technologies that could support non-intrusive eye and gaze tracking in a museum context.

I. Problems associated with gaze tracking input as an interaction mechanism

The main problem associated with use of eye movements and gaze direction as an interaction mechanism is known in the literature as "Midas touch" or "the clutch" problem (Jacob, 1993). In simple terms, the problem is that if looking at something should trigger an action, one would be triggering this action even by just observing a particular element on the display (or projection). The problem has been addressed numerous times in literature, and there are many proposed technical solutions. Detailed analysis and overview of these solutions is beyond the scope of this paper. I will present here only a few illustrative examples.

One of the solutions to the *Midas Touch* problem, one developed by Risø National Research Laboratory, was to separate the gaze-responsive area from the observed object. The switch (aptly named EyeCon) is a square button placed next to the object that one wants to interact with. When the button is focused (ordinarily for half a second), it 'acknowledges' the viewer's intent to interact with an animated sequence depicting a gradually closing eye. The completely closed eye is equivalent to the pressing of a button (see Figure 3).

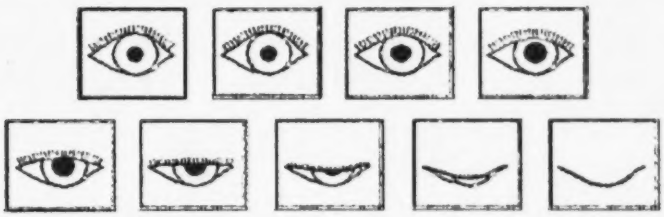

Figure 3. An EyeCon activation sequence. Separating the control mechanism from interactive objects allows natural observation of the object (image reproduced from Glenstrup, A.J., Engell-Nielsen, T., 1995)

One of the problems with this technique comes from the very solution – it is the *separation of selection and action*. The other problem is the *interruption of the flow of interaction* – in order to select (interact with) an object, the user has to focus on the action button for a period of time. This undermines the unique quality of gaze direction as the fastest and natural way of pointing and selection (focus).

Another solution to the same problem (with very promising results) was to provide the 'clutch' for interaction through another modality - voice (Glenn, Iavecchia, Ross, Stokes, Weiland, Weiss, Zakland 1986) or manual (Zhai, Morimoto, Ihde 1999) input.

The second major problem with eye movement input is the sheer volume of data collected during eye-tracking and its meaningful analysis. Since individual fixations carry very little meaning on their own, a wide range of eye tracking metrics has been developed in the past 50 years. An excellent and very detailed overview of these metrics can be found in Jacob (2003/in print). Here, I will mention only a few that may be used to infer viewer's interest or intent:

- *number of fixations*: a concentration of a large number of fixations in a certain area may be related to a user's interest in the object or detail presented in that area when viewing a scene (or a painting). Repeated, retrograde fixations on a certain word while reading text are taken to be indicators of increased processing load (Just, Carpenter 1976).
- *gaze duration*: gaze is defined as a number of *consecutive* fixations in an area of interest. Gaze duration is the total of fixation durations in a particular area.
- *number of gazes*: this is probably a more meaningful metric than the number of fixations. Combined with gaze duration, it may be indicative of a viewer's interest.
- *scan path*: the scan path is a line connecting consecutive fixations (see Figure 2, for example). It can be revealing of a viewer's visual exploration strategies and is often very different in experts and novices.

The problem of finding the right metric for interpretation of eye movements in a gallery/museum setting is more difficult than in a conventional research setting because of the complexity of the visual stimuli and the wide individual differences of users. However, the problem may be made easier to solve by dramatically constraining the number of interactions offered by a particular application and making them correspond to the user's expectations. For example, one of the applications of the interface I will propose is a simple gaze-based browsing mechanism that allows the viewer to quickly and effortlessly leaf through a museum collection (even if he/she is a quadriplegic and has retained only the ability to move the eyes).

II. Gaze-based interface for museum content

Needless to say, even a gaze-based interface that is specifically designed for museum use has to provide a solution for general problems associated with the use of eye movement-based interactions. I will approach this issue by analyzing three different strategies that may lead to the solution of the *Midas touch* problem. These strategies differ in terms of the of the interaction mechanism, as it relates to:

- time
- location, and
- user action

It is clear that *any* interaction involves time, space and actions, so the above classification should be taken to refer to the key component of the interface solution. Each of these solutions has to accommodate two modes of operation:

- the observation mode, and
- the action (command) mode

The viewer should have a clear indication as to which mode is currently active, and the interaction mechanism should provide a way to switch between the modes quickly and effortlessly.

Time-based interfaces

At first glance, a time-based interface seems like a good choice (evident even for myself when choosing the title of this paper). An ideal setup (for which I will provide more details in the following sections) for this type of interface would be a high-resolution *projection* of a painting on the screen with an eye-tracking system concealed in a small barrier in front of the user. An illustration of a time-based interaction mechanism is provided in Figure 4. The gaze location is indicated by a traditional cursor as long as it remains in a non-active (in this case, outside of the painting) area. When the user shifts the gaze to the gaze-sensitive object (painting), the cursor changes its shape to a faint circle, indicating that the observed object is aware of the user's *attention*. I have chosen the circle shape because it does not interfere with the viewer's observation, even though it clearly indicates potential interaction. As long as the viewer continues visual exploration of the painting there is no change in status. However, if the viewer decides to focus on a certain area for a predetermined period of time (600 ms), the cursor/circle starts to shrink (zoom), indicating the beginning of the *focusing* procedure.

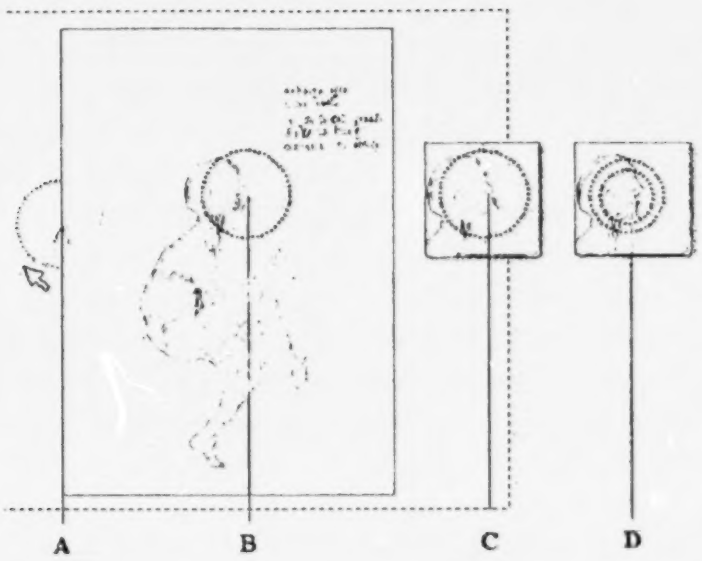

Figure 4. The cursor changes at position (A) into focus area indicating that the object is 'hot'.

Position (B) marks the period of relative immobility of the cursor and the beginning of the focusing procedure. Relative change in the size of the focus area (C) indicates that focusing is taking place. The appearance of concentric circles at time (D) indicates imminent action. The viewer can exit the focusing sequence at any time by moving the point of observation outside of the current focus area.

If the viewer continues to fixate on the area of interest, the focusing procedure continues for the next 400 milliseconds, ending with a 200 millisecond long signal of *imminent action*. At any time during the focusing sequence (including the imminent action signal), the viewer can return to observation mode by moving the gaze away from the current fixation point. In the scenario depicted above (and in general, for time-based interactions) it is desirable to have *only one* pre-specified action relevant to the context of viewing. For example, the action can be that of zooming-in to the observed detail of the painting (see Figure 6), or proceeding to the next item in the museum collection. The drawbacks of time-based interaction solutions triggered by focusing on the object/area of interest areas follows:

- the problem of *going back* to observation mode. This means that the action triggered by focusing on a certain area has to be either *self-terminating* (as is the case with the 'display the next artifact' action, where the application switches automatically back to the observation mode) ,or one has to provide a simple mechanism that would allow the viewer to return to the observation mode (for example, by moving the gaze focus outside of the object boundary);
- the problem of choice between *multiple actions*. Using the time-based mechanism, it is possible to trigger different actions. By changing the cursor/focus shape, one can also indicate to the viewer which action is going to take place. However, since the actions are tied to the objects themselves, the viewer essentially has no choice but to accept the pre-specified action. This may not be a problem in a context where pre-specified actions are meaningful and correspond to the viewer's expectations. However, it does limit the number of actions one can 'pack' into an application and can create confusion in cases where two instances of focusing on the same object may trigger off different actions.
- the problem of *interrupted flow* or *waiting*. Inherent to time-based solutions is the problem that the viewer always has to wait for an action to be executed. In my experience, after getting acquainted with the interaction mechanism, the waiting time becomes subjectively longer (because the users know what to

BEST COPY AVAILABLE

expect) and often leads to frustration. The problem can be diminished to some extent by progressively shortening the duration of focusing necessary to trigger the action. However, at some point it can lead to another source of frustration since the viewer may be forced to constantly shift the gaze around in order to stay in the observation mode.

Inspite of the above mentioned problems, time-based gaze interactions can be an effective solution for museum use where longer observation of an area of interest provides the viewer with more information. Another useful approach is to use the gaze direction as input for the delivery of additional information through another modality. In this case, the viewer does not need to get visual feedback related to his/her eye movements (which can be distracting on its own). Instead, focusing to an area of interest may trigger voice narration related to viewer's interest. For an example of this technique in the creation of a gaze-guided interactive narrative, see Starker & Bolt (1990).

Location-based interfaces

Another traditional way of solving the "clutch" problem in gaze-based interfaces is by separating the modes of observation and action by using controls that are in the proximity of the area of interest but do not interfere with visual inspection. I have already described EyeCons (Figure 3) designed by the Risø National Research Laboratory in Denmark (for a detailed description see Glenstrup and Engell-Nielsen, 1995). In the following section I will first expand on EyeCons design and then propose another location-based interaction mechanism. The first approach is illustrated in Figure 5.

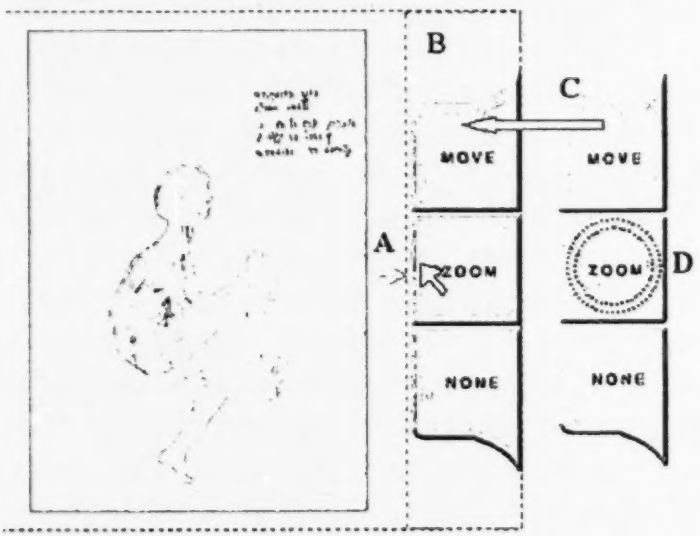

Figure 5. Movement of the cursor (A) into the gaze-sensitive area (B) slides into view the action palette (C).

Fixating any of the buttons is equivalent to a button press and chooses the specified action which is executed without delay when the gaze returns to the object of interest. The viewer can also return to observation mode by choosing no action button. The action palette slides out of view as soon as the gaze moves out of the area (B).

The observation area (the drawing) and the controls (buttons) are separated. At first glance, the design seems very similar to that of the EyeCons, but there are some enhancements that make the interactions more efficient. First, the controls (buttons) are located on a configurable 'sliding palette', a mechanism that was adopted by the most widely used operating system (Windows) in order to provide users with more

BEST COPY AVAILABLE

'screen real estate'. The reason for doing this in a museum context is also to minimize the level of distraction while observing the artifact. Shifting the gaze to the side of the projection space (B) slides the action palette into the view. The button that is currently focused becomes *immediately* active (D) signaling the change of mode by displaying the focus ring and changing the color. This is a significant difference compared to the EyeCons design, which combines both location- and time-based mechanisms to initiate action. Moving the gaze back to the object leads to the execution of specified action (selection, moving, etc.). Figure 6 illustrates the outcome of choosing the 'zoom' action from the palette. The eye-guided cursor becomes a magnifying glass allowing close inspection of the artifact.

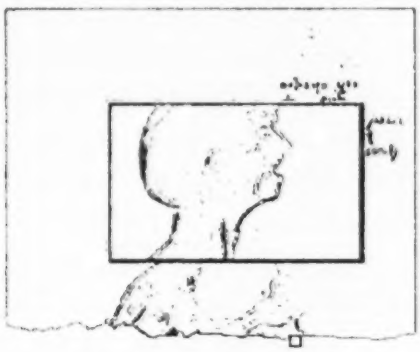

Figure 6. After choosing the desired action (see Figure 5), returning the gaze to the object executes the action without delay. The detail above shows the 'zoom-in' tool, which becomes 'tied' to the viewer's gaze and allows close inspection of the artifact.

One can conceptually expand location-based interactions by introducing the concept of an *active surface*. Buttons can be viewed as being essentially single-action locations (switches). It really does not matter which part of the button one is focusing on (or physically pressing) – the outcome is always the same. In contrast, a surface affords assigning meaning to a series of locations (fixations) and makes possible *incremental* manipulation of an object.

Figure 7 provides an example of a surface-based interaction mechanism. Interactive surfaces are discretely marked on the area surrounding the object. For the purpose of illustration, a viewer's scan path (A) is shown superimposed over the object and indicates gaze movement towards the interactive surface. Entering the active area is marked by the appearance of a cursor in a shape that is indicative of the possible action (D). The appearance of the cursor is followed by a brief latency period (200-300 ms) during which the viewer can return to the observation mode by moving the gaze outside of the active area. If the focus remains in the active area (see Figure 8), any movement of the cursor along the longest axis of the area will be *incrementally mapped* onto an action sequence – in this case, rotation of the object.

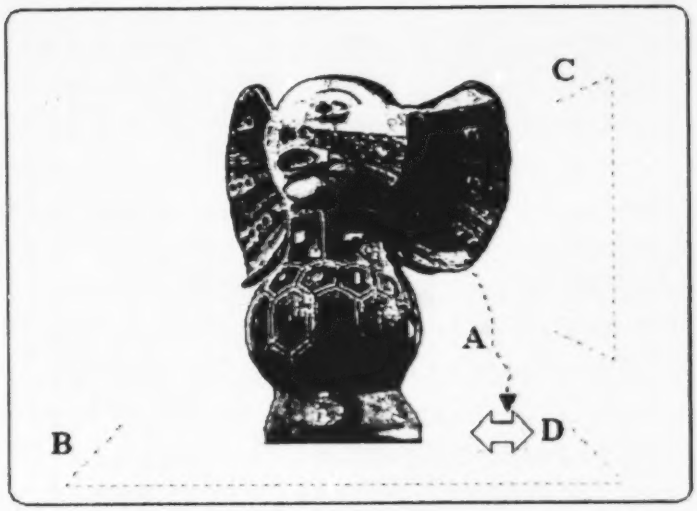

Figure 7. Surface-based interaction mechanism. Viewer's scanpath is visible at (A). Two interactive surfaces (B and C) are discretely marked on the projection. Moving the gaze into the area of interactive surface is marked by appearance of cursor with the shape indicative of possible action (D).

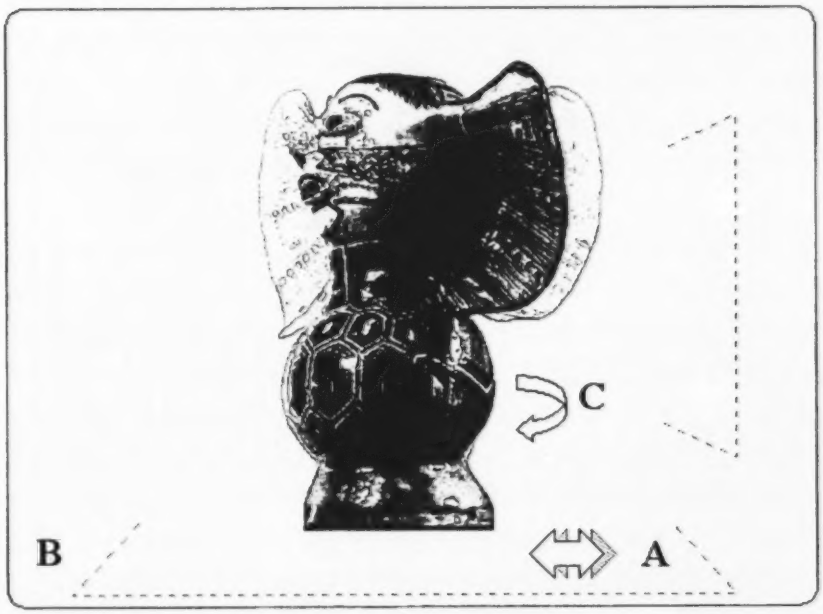

Figure 8. If the viewer's gaze (as indicated by cursor position at A) remains within interactive surface (B), any gaze movement within the surface will lead to incremental action – in this case rotation of the object (C).

The advantages of surface-based interaction mechanisms are the introduction of more complex, incremental action sequences into eye movement input and the possibility of rapid shifts between the observation and action modes. The drawback is

that the number of actions is limited and that the surfaces, although visually non-intrusive, still claim a substantial portion of the display.

Action-based interfaces

Building on the previous two models, one can further expand the conceptual framework for gaze-based interfaces. This time I will focus on the *gaze action* as a mechanism for switching between the observation and the active (command) mode. Analysis of the previously described surface-based model reveals that it can be described as an intermediary step between the surface- and action-based interfaces. In this model, although the shift between the observation and action mode is dependent on *the location* of gaze focus, the control of interaction is based on gaze action (moving the focus/cursor over gaze-sensitive surface). Thus, the last step in our analysis is to explore the possibility of using predominantly *gaze-based actions* as a control mechanism. This may seem like slippery ground because physiologically our visual behavior is mostly geared towards collecting information and not acting upon the world. The exception of a kind is in the domain of sexual and social behaviors where gaze direction and duration may literally have physical consequences by signaling attraction, dominance, submissiveness, etc. Fine literature abounds with examples describing gazes as having a tangible effect ("his piercing gaze," "he felt her gaze boring two little holes at the back of his neck...," "her angry gaze was whipping across the room trying to find out who did this to her.." to mention a few). Our ability to transfer knowledge from one sensory domain to another modality will be the key component in the proposed outline of an action-based gaze interface.

In eye-tracking literature, a gaze is most often defined as a number of consecutive fixations in a certain area. This metric emphasizes the *location* and the *duration* characteristics of the gaze and can be extremely useful in inferring the viewer's interest or gauging the complexity of the stimulus. However, in my proposal I would like to focus on two, often neglected, characteristics of a *moving gaze* that can be consciously used by a viewer to indicate his/her *intention*. These are:

- the direction of gaze movement, and
- the speed of gaze movement.

For technical purposes a moving gaze can be defined as a number of consecutive fixations progressing in the same direction. It corresponds roughly to longer, straight parts of a scan path and is occasionally referred to as a *sweep* (Altonen et al. 1998). The reason for choosing these characteristics is twofold. First, eyes can move much faster than the hand (and there is evidence from literature that eye-pointing is significantly faster than mouse pointing, see Sibert and Jacob 2000). Second, as mentioned before, directional gaze movement is often used in social communication. For example, we often indicate in a conversation exactly 'who' we are talking about by repeatedly shifting the gaze in the direction of the person in question.

In order to create an efficient gaze-based interface, one has to be able to replicate the basic mouse-based actions used in the traditional graphical user interface (GUI). These are: **pointing** (cursor over), **selection** (mouse down), **dragging** (mouse down + move) and **dropping** (mouse up). I will also propose the inclusion of yet another non-traditional action, which I introduced in interface design a while ago (Milekic, 2000) and which proved to work extremely well as an intuitive browsing mechanism. This is the action of *throwing* which is dependent on the speed of movement of a selected object. Compared to the traditional interface, the throwing action is an expansion of the action of dragging an object. As long as the speed of dragging remains within a certain limit, one can move an object anywhere on the screen and drop it at desired location. However, if one 'flicks' the object in any direction, the object is released and literally 'flies away' (most often, to be replaced by another object). I have implemented this mechanism in a variety of mouse-, touchscreen- and gesture-based installations in museums and it has been successfully used by widely diverse audiences, including very young children. Subjectively, the action is very intuitive and natural, and the feeling can be best compared to that of sliding a glass on a polished surface (a skill that many bar tenders hone to perfection). In the following sections I will describe each of the gaze-based actions.

Gaze-pointing (Figure 9) is the easiest function to replicate in a gaze-based interface. It essentially consists of a visual clue that indicates to the viewer which area of the display is currently observed. Although one can use the traditional cursor for this purpose, it is desirable to design a cursor that will not interfere with observation. Dynamic change of cursor shape when moving over different objects can also be used to indicate whether an object is gaze-sensitive and to specify the type of action one can initiate (this technique is used in surface-based interface, described above; see Figure 4, for example). I have chosen a simple dashed circle as an indicator of the current gaze location. Pointing action is maintained as long as there are no sudden substantial changes in a specific gaze direction. If such a change occurs, the tracking algorithm determines the *direction* of gaze movement and, if necessary, initiates appropriate action.

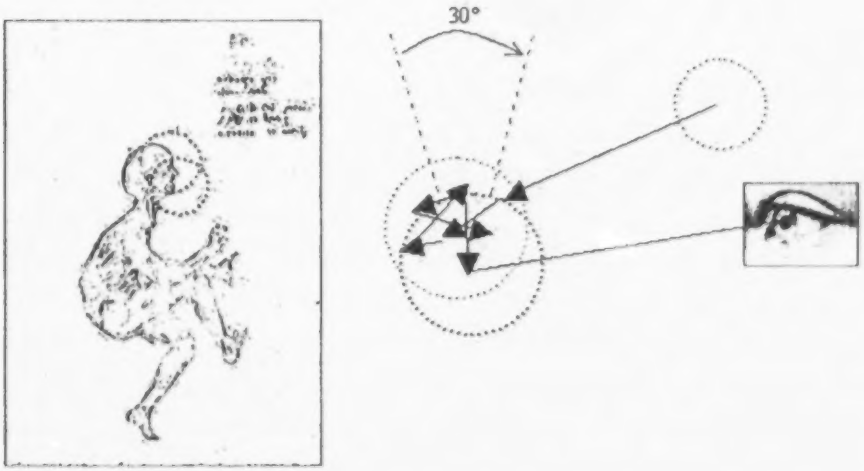

Figure 9. Gaze-pointing. The viewer can observe the artifact with the pointing cursor (dashed circle) indicating the current gaze location. Sweeping gazes across the scene are possible as long as they are not in upward direction and end in the 30° angle strip.

This does not mean that the viewer is limited to slow (and unnatural) observation. In fact, switching from observation to action mode (selection) occurs *only* if movement of sufficient amplitude occurs in an upward direction and ends up in a fairly narrow area spanning approximately 30° above the current focus area. This means that viewers can, more or less, maintain a normal observation pattern, even if it includes sweeping gaze shifts, as long as they don't end up in the critical area.

Gaze-selection (Figure 10) is an action initiated by a sudden upward gaze shift. The action is best described (and subjectively feels like) the act of upward stabbing, or 'hooking' of the object. In a mouse-based interface the selection is a separate action – that is, one can just select an object, or select-drag-drop it somewhere else, or de-select it. In a gaze-based interface, what happens after the selection of an object will depend on the context of viewing. When multiple objects are displayed, the selection mechanism can act as a self-terminating action, making it possible for the viewer to select a subset of objects. In this case, highlighting the object would indicate the selection. However, in the museum context (assuming that the viewers will most often engage in observation of a single artifact) object selection may just be a prelude to the action of moving (dragging). In this case the object becomes, figuratively speaking, 'hooked' to the end of the viewer's gaze, as indicated by a change of the cursor's shape to that of a target.

BEST COPY AVAILABLE

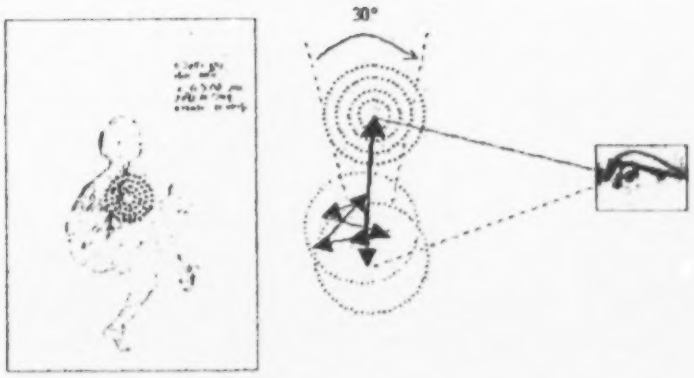

Figure 10. Gaze-selection. Shifting the gaze rapidly upwards within the 30° triggers of the selection process. The cursor changes the shape to that of a target and positions itself at the center of the object as a prelude to the action of gaze-dragging.

Gaze-dragging (Figure 12). Once the object has been selected ('hooked' to the viewer's gaze), it will follow the viewer's gaze until it is 'dropped' at another location. This action is meaningful in cases when the activity involves the repositioning of multiple objects (for example, assembling a puzzle). In the scenario depicted above, the viewer can 'throw away' the current object and get a new one.

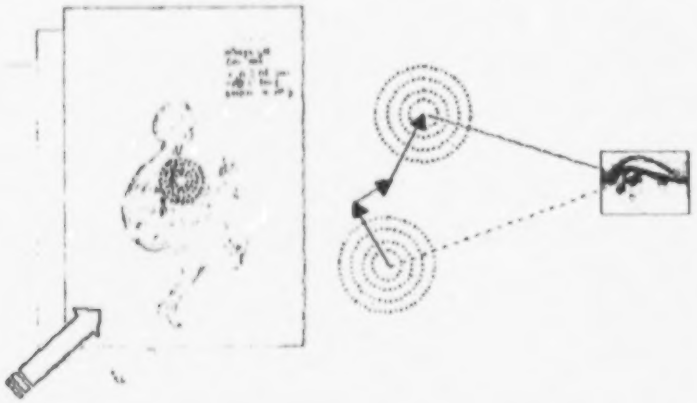

Figure 11. Gaze-dragging. The painting is 'hooked' to viewer's gaze and follows its direction. At this stage the viewer can decide either to 'drop' the painting at another location (see Figure 12) or, 'throw' away the current one and get a new artifact.

BEST COPY AVAILABLE

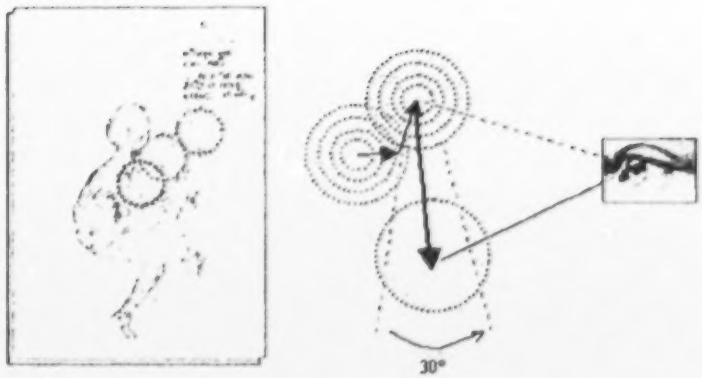

Figure 12. Gaze-dropping. The action of dropping an object is the opposite of 'hooking' it. A quick downward gaze movement releases the object and switches the application into observation mode.

Gaze-throwing (Figure 13) is a new interaction mechanism that allows efficient browsing of visual data bases with a variety of input devices, including gaze input. An object that has been previously selected ("hooked") will follow the viewer's gaze as long as the speed of movement does not exceed a certain threshold. A quick glance to the left or the right will release the object and it will 'fly away' from the display to be replaced by a new artifact.

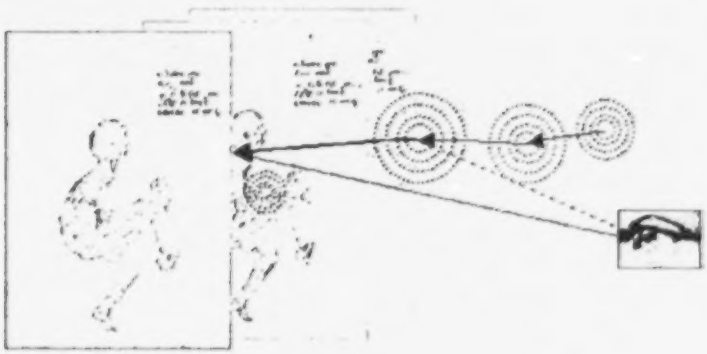

Figure 13. Gaze-throwing. 'Throwing' an object away is accomplished by moving the gaze rapidly to the left or to the right. Once the object reaches threshold speed it is released and 'flies away'. A new artifact floats to the center of display.

The objects appear in a sequential order, so if a viewer accidentally throws an object away, it can be recovered by throwing the next object in the opposite direction.

To summarize, action-based gaze input mechanisms have the advantage of allowing the viewer to *act upon the object* at will, without time or location constraints. The mechanism is simple and intuitive because it is analogous to natural actions in other modalities. The best way to think about action-based gaze input is as a kind of *eye-graffiti*. The vocabulary of suggested gaze-gestures for eye input is presented in Figure 14. It is similar to the text input mechanism used for Palm personal organizers where the letters of the alphabet are reduced to corresponding simplified gestures. The fact that millions of users were able to adopt this quick and efficient text input mechanism is an indication that the development of *eye-graffiti* has significant potential for gaze based interfaces.

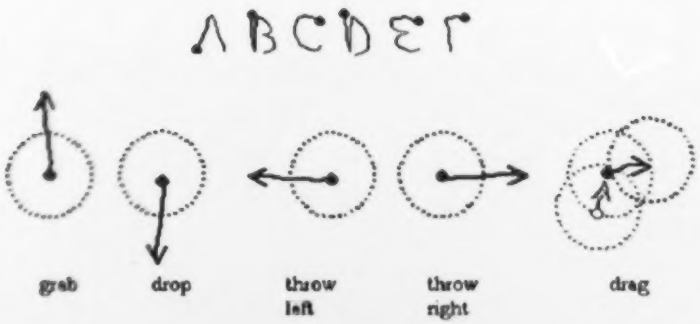

Figure 14. Eye-graffiti. Top row presents graffiti used for text input (letters A,B,C,D,E,F respectively) in Palm OS based personal organizers. Bottom row outlines suggested gaze-gestures that trigger different actions once the object has been selected.

The dashed circle in the illustration above does not represent the visual representation of the cursor, but rather the area used to calculate the direction and the velocity of gaze movement by the tracking algorithm. The heavy dot indicates the starting point of a gesture. However, while action-based gaze input mechanism may seem best suited for museum applications, the ideal interface is probably a measured combination of all three approaches.

III. Current Technologies for Non-Intrusive Eye Tracking

Unlike in the laboratory environments, the eye-tracking technology used in a museum setting has to meet additional specific requirements. Some of the most obvious ones are:

- it should be non-intrusive. This excludes all eye-tracking devices that use goggles, head-straps, chin-rests or such.
- it should allow natural head movements that occur during viewing.
- it should not require individual calibration.
- it should be able to perform with a wide variety of eye shapes, contact lenses or glasses.
- it should be portable.
- it should be affordable.

With the ncreasing processor speeds of currently available personal computers, it seems that the most promising eye-tracking technology is that based on digital video analysis of eye movements. The most commonly used approach in video-based eye tracking is to calculate the angle of the visual axis (and the location of the fixation point on the display surface) by tracking the relative position of the pupil and a speck of light reflected from the cornea, technically known as the "glint" (see Figure 15). The accuracy of the system can be further enhanced by illuminating the eye(s) with low-level infra-red lightto produce the "bright pupil" effect and make the video image easier to process (B in Figure 15). Infrared light is harmless and invisible to the user.

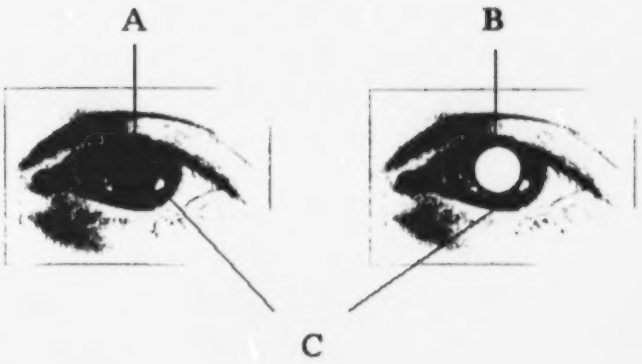

Figure 15. Gaze direction can be calculated by comparing the relative position and the relationship between the pupil (A) and corneal reflection – the glint (C). Infra-red illumination of the eye produces the 'bright pupil' effect (B) and makes the tracking easier.

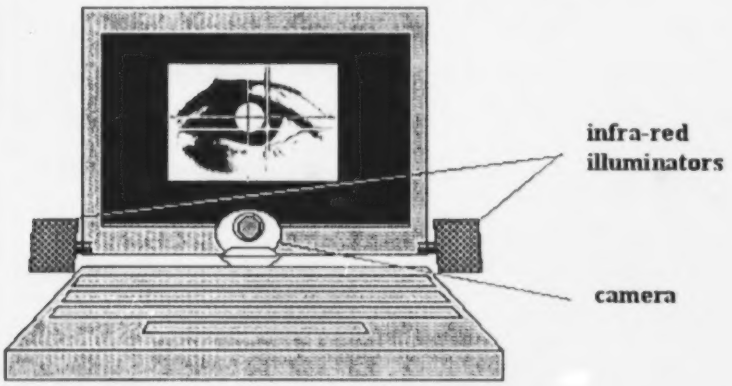

Figure 16. Several manufacturers produce portable eye-tracking systems similar to the one depicted above. While the camera position is most often bellow the eye level (eyelids interfere with tracking from above), the shape and position of infrared illuminators vary from manufacturer to manufacturer.

A typical and portable eye-tracking system similar to the ones commercially available is depicted in Figure 16. Since the purpose of this paper is not to endorse any particular manufacturer, I urge interested readers to consult the large Eye Movement Equipment Database (EMED) available on the World Wide Web. Keeping in mind that many museums and galleries have very modest budgets, I will specifically address the issue of *affordable* eye-tracking systems.

The price range of most commercially available eye-trackers is between $5000 and $60.000, often with additional costs for custom software development, setup etc. Although there are some exceptions, the quality and the precision of the system tend to correlate with the price. However, with the increasing speed of computer processors, greater availability of cheap digital video cameras (like the ones used for Web-based video conferencing) and, most importantly, the development of sophisticated software for video signal analysis, it is becoming possible to build eye-trackers within a price range comparable to that of a new personal computer. Even though the cheaper systems have lower spatial and temporal resolution when compared to the research equipment, in a museum/gallery setting they may be used for different applications; for example, for browsing a museum collection with additional information provided by voice-overs. A more significant use would be

providing access to the museum content to visitors with special needs. An example of a cost-effective solution based on a personal computer and a Web-cam for eye-gaze assistive technology was recently described (Corno, Farinetti and Signorile, 2002).

Most commercially available eye-tracking systems (including the high-end ones) have two characteristics that make them less than ideal for use in museums. These are:

- the system has to be calibrated for each individual user
- even remote eye-trackers have very low tolerance for head movements and require the viewer to hold the head unnaturally still, or to use external support like head- or chin-rests.

The solution lies in the development of software able to perform eye-tracking data analysis in more natural viewing circumstances. A recent report by Quiang and Zhiwei (2002) seems to be a step in the right direction. Instead of using conventional approaches to gaze calibration, they introduced a procedure based on neural networks that incorporates natural head movements into gaze estimation and eliminates the need for individual calibration.

The emergence of eye-tracking technologies based on a personal computer equipped with a Web-cam and the development of software that allows gaze tracking in natural circumstances open up a whole new area for museum applications. The described technologies make Web-based delivery of gaze-sensitive applications possible. This not only presents an opportunity for a novel method of content delivery (and reaching different groups of users with special needs) but also offers an incredible possibility to collect, on a massive scale, data related to visual analysis of museum artifacts. However, a word of caution is in order here. One cannot overemphasize the importance of *context* in an eye-tracking application (or, for that matter, in any application). In an appropriate context, even a fairly simple setup can produce magical results, and the use of the most expensive equipment can lead to viewer frustration in a flawed application.

Conclusion

I have outlined a conceptual framework for the development of a gaze-based interface for use in a museum context. The major component of this interface is the introduction of ***gaze gestures*** as a mechanism for performing intentional actions on observed objects. In conjunction, an overview of suitable eye-tracking technologies was presented with an emphasis on low cost solutions. The proposed mechanism allows the development of novel and creative ways for content delivery both in a museum setting and via the World Wide Web. An important benefit of this approach is that it makes museum *content* (and not just the building or the restrooms) accessible to a wide variety of populations with special needs. It also offers the possibility of data-logging related to visual observation on a massive scale. These records can be used to further refine the content delivery mechanism and to promote our understanding of both the psychological and the neurophysiological underpinnings of our relationship with the Art.

References

Altonen, A., A.Hyrskykari, K. Raiha. (1998). 101 Spots, or how do users read menus? in Proceedings of CHI 98 Human Factors in Computing Systems, ACM Press, pp 132-139.

Bever, T.G., (1970). The cognitive basis for linguistic structure, in J.R. Hayes, ed., *Cognitive development of language*, Wiley, New York.

Card, S.K. (1984). Visual search of computer command menus, in H. Bouma, D.G. Bouwhuis (Eds.) *Attention and Performance X, Control of Language Processes*, Hillsdale, NJ, LEA.

Corno, F., L.Farinetti, I. Signorile (2002). A Cost-Effective Solution for Eye-Gaze

Assistive Technology, ICME 2002: IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland.

Corkum V., Moore C. (1998). The origins of joint visual attention in infants. Developmental Psychology, 34, pp 28–38.

Dodge, R., T.S. Cline (1901). The angle velocity of eye-movements. Psychological Review, 8, 145–57.

Duchowski, A.T. (2002). Eye Tracking Methodology: Theory and Practice, Springer Verlag.

Glenn III, F.A., H.P.Iavecchia, L.V.Ross, J.M. Stokes, W.J. Weiland, D. Weiss, A.L. Zakland, (1986). Eye-voice-controlled interface, Proceedings of the Human Factors Society, 322-326.

Glenstrup, A.J., T. Engell-Nielsen, (1995). Eye Controlled Media: Present and Future State. Minor Subject Thesis, DIKU, University of Copenhagen, available at: http://www.diku.dk/~panic/eyegaze/article.html#contents.

Hartridge, H., L.C. Thompson, (1948). Methods of investigating eye movements, British Journal of Ophthalmology, 32, pp 581-591.

Hutchinson, T.E., K.P. White, W.N. Martin, K.C. Reichert, L.A. Frey, (1989). Human-Computer Interaction Using Eye-Gaze Input, IEEE Transactions on Systems, Man, and Cybernetics, 19, pp 1527-1534.

Jacob, R. J. K. (1993). Eye-movement-based human-computer interaction techniques: Toward non-command interfaces, in H. R. Hartson & D. Hix, (eds.) *Advances in Human-Computer Interaction*, Vol. 4, pp 151-190, Ablex Publishing Corporation, Norwood, New Jersey.

Jacob, R.J.K., K.S. Karn, (2003). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), in *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*, J. Hyona, R. Radach, H. Deubel (Eds.), Oxford, Elsevier Science (in press).

Judd, C.H., C.N. McAllister, W.M. Steel, (1905). General introduction to a series of studies of eye movements by means of kinetoscopic photographs, in J.M. Baldwin, H.C. Warren & C.H. Judd (Eds.) Psychological Review, Monograph Supplements, 7, pp1-16, The Review Publishing Company, Baltimore.

Just, M.A., P.A. Carpenter, (1976). The role of eye-fixation research in cognitive psychology, Behavior Research Methods & Instrumentation, 8, pp 139-143.

Levine, J.L. (1981) An Eye-Controlled Computer, Research Report RC-8857, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y.

Milekic, S. (2000) Designing Digital Environments for Art Education / Exploration, Journal of the American Society for Information Science, Vol. 51-1, 49-56, Wiley.

Moore, C. (1999) Gaze following and the control of attention, in P. Rochat (ed.), *Early social cognition: understanding others in the first months of life*, pp 241–256.

National Gallery, London "Telling Time Exhibition", Web source consulted 1/24/03, available at: http://ibs.derby.ac.uk/gallery/updates.shtml

Scaife M., J. S. Bruner (1975). The capacity for joint visual attention in the infant. Nature, 253, pp 265–266.

Sibert, L. E., R.J.K. Jacob (2000). Evaluation of Eye Gaze Interaction, Proceedings of the CHI 2000, ACM, New York, 281-288, available at: http://citeseer.nj.nec.com/article/sibert00evaluation.html

Starker, I., R.A. Bolt, (1990). A gaze-responsive self-disclosing display, in *CHI '90 Proceedings*, ACM, pp. 3-9.

Stewart, J., & C. Logan, (1998). *Together: Communicating Interpersonally (5th Ed.)*. Boston: McGraw Hill, 90-100.

Qiang, J., Z. Zhiwei, (2002). Eye and Gaze Tracking for Interactive Graphic Display, International Symposium on Smart Graphics, June 11-13, 2002, Hawthorne, NYVertegaal, R. (1999). The GAZE groupware system: mediating joint attention in multiparty communication and collaboration, in *Proceedings of the ACM CHI'99 Human Factors in Computing Systems*, ACM Press, New York, pp 294-301.

Ware,C., H.H. Mikaelian, (1987). An evaluation of an eye tracker as a device for computer input, Proceedings of the CHI 1987, ACM, New York, 183-188.

Yarbus, A.L. (1967) Eye movements during perception of complex objects, in L. A. Riggs, (ed.), *Eye Movements and Vision*, Plenum Press, New York, 171-196.

Zhai, S., C. Morimoto, S. Ihde, (1999). Manual and Gaze Input Cascaded (MAGIC) Pointing, Proceedings of the CHI'87, ACM New York, 246-253.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").